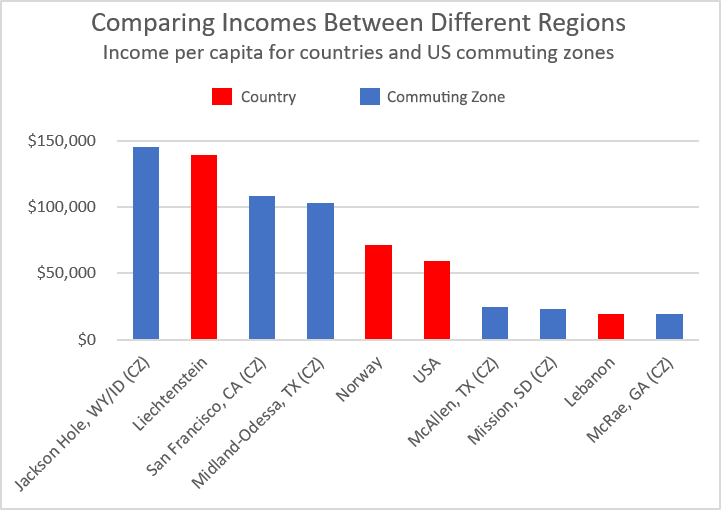
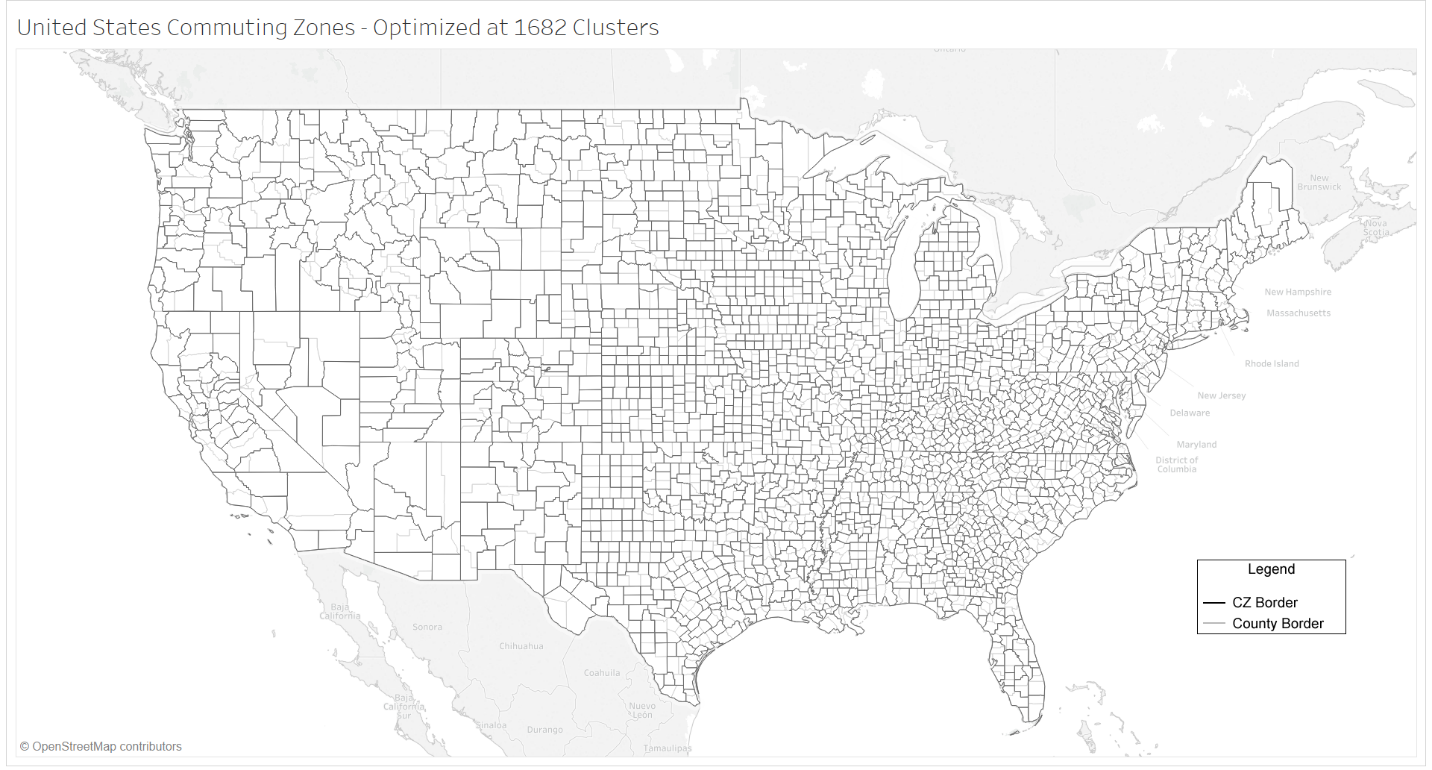
The 1682 States of the USA

We begin most discussions of the US economy by making a big mistake. The US is not one monolithic region. New York City might have more in common with London, England than Paris, Texas. San Francisco has a per capita income five times more than regions in rural America. Different regions experience different realities, and policies related to trade, minimum wage, and other issues create dramatically different regional impacts.

And yet, when we try to split the US into more accurate economic regions, we run into a Goldilocks dilemma: too big, too small, or not comprehensive enough. Regions like states and the Economic Areas (from the Bureau of Economic Analysis) are too big. They do a good job of covering the entire country, but combine too many diverse regions. In contrast, counties are too small. Large economies expand across multiple counties. And while Metropolitan Statistical Areas do a better job of grouping similar counties together, they do not cover much of the country. The solution to this dilemma is Commuting Zones (CZs), but prior methods for defining these regions are outdated and un-transparent.

If we think about the economy as the collection of problems a group of people is paid to solve, CZs do a great job of defining the “group”. Using a clustering algorithm, this methodology finds discreet groups of counties where the residents work together. As compared to previous versions of CZs, we can also add three improvements due to improved computing power and modern practices. First, we can make the analysis more robust by analyzing the entire country at once, rather than breaking the data into regions that must be pieced together post-hoc. Second, we can make it more accurate by finding the optimal number of clusters - or CZs - rather than relying upon a rule of thumb. Third, we can make it more accessible and improvable by making the code open source.

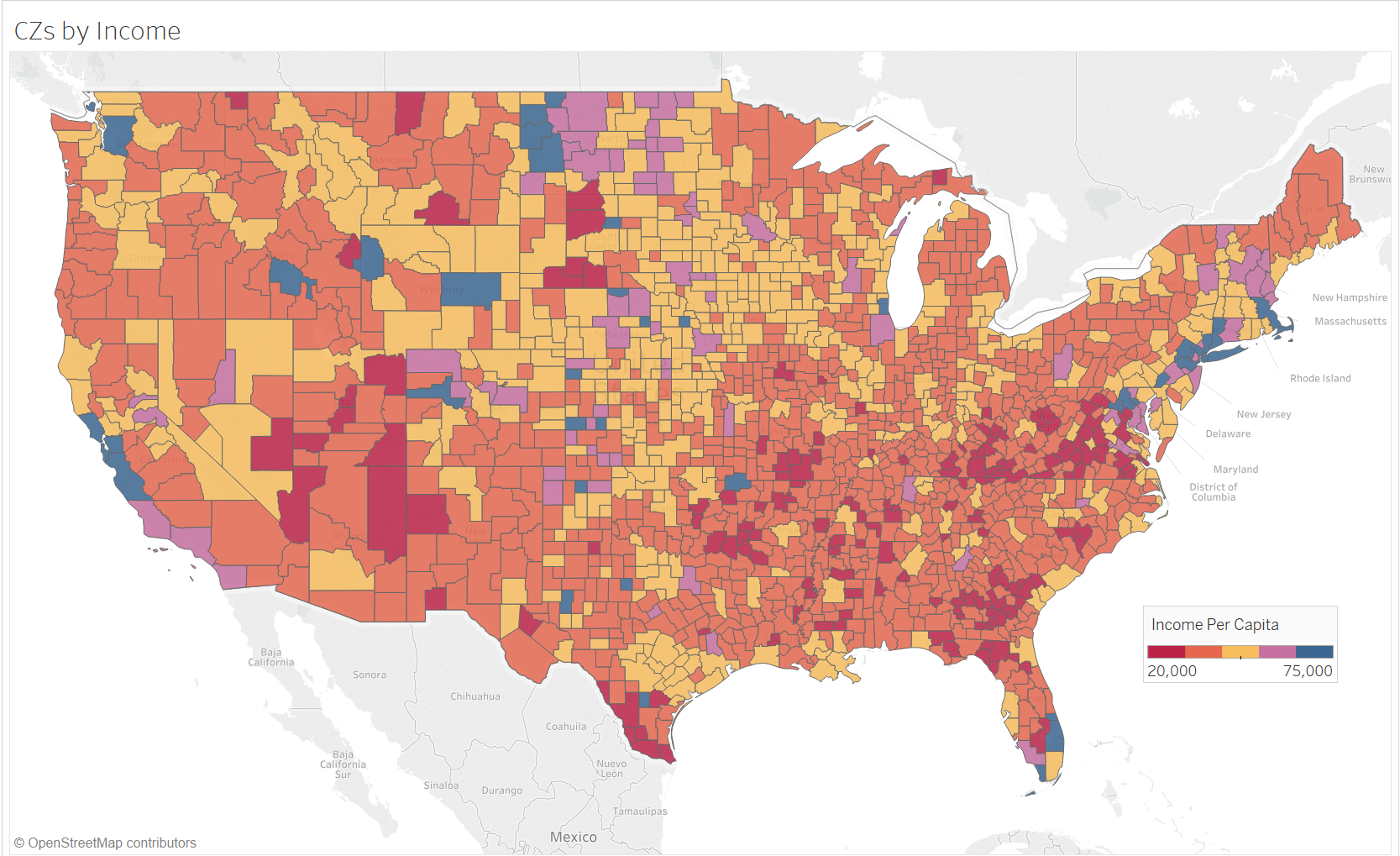
As a result, we get 1682 different CZs, representing the groups of people within the US who work together. Yes, 1682 seems like a large number. It’s still much lower than the 3221 counties, parishes, and provinces within US states and territories. Also, many isolated counties remain isolated, and these 672 CZs with only one county substantially increase the total number.



Some economic development practitioners, such as chambers of commerce, might prefer existing practices, such as using a 60-minute driving radius to determine their labor area. For business recruitment, using drive-times might make sense, but we can still create value by considering the status quo represented by CZs. On the other hand, for talent development and business retention, CZs help reveal “what you’re working with”: the people within these counties provide the workforce for local business, and the businesses within these counties create the local economy.

My goal is not to claim that 1682 is the right number of clusters, or that I created the perfect method. I have three goals. First, I want to facilitate a process for creating a better method – we should never rely on the current status quo that uses un-transparent code. Second, I want to improve awareness about the many different economies that exist within our national borders. Most importantly, I want to better understand the different realities that regions face.

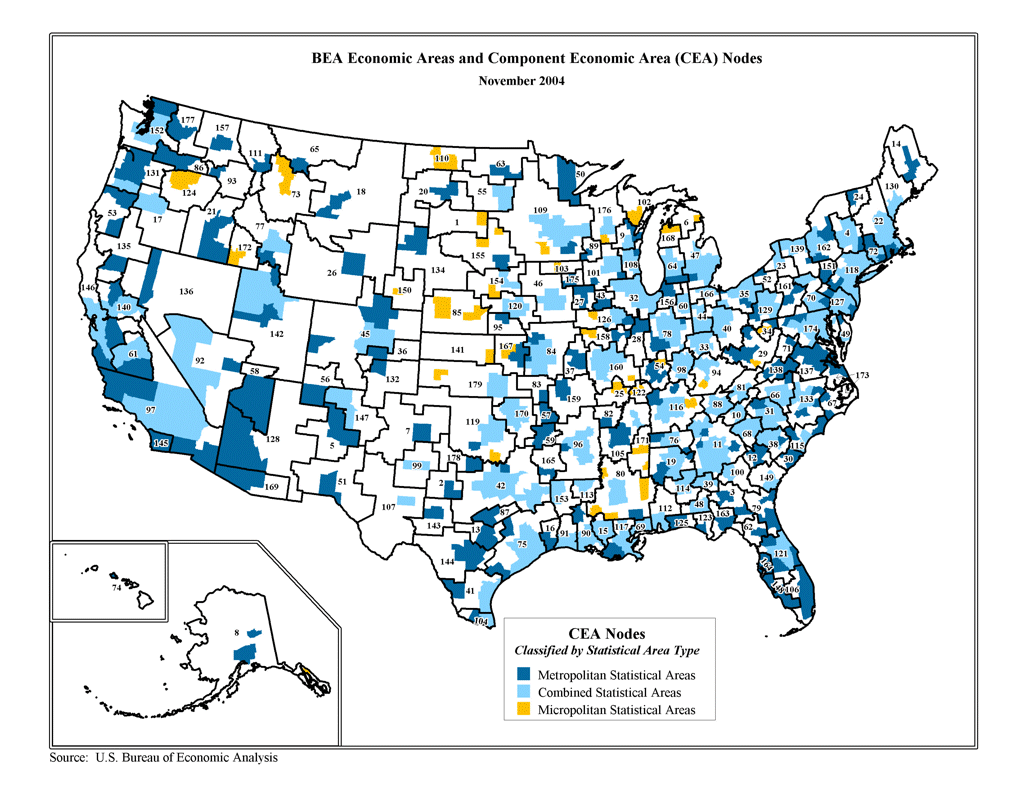
The map below not only shows the differences between our regions – in terms of income per person within CZs - it should also discourage us from talking about the national economy as one big thing. For example, a high minimum wage in the blue areas makes more sense than in lower-income orange areas. International trade might create net gains on the national level, but the richer blue areas with high-value services often gain jobs while the less rich red areas that rely on manufacturing lose jobs. These economic issues are complicated. In order to do a better job talking about them, we need a better common understanding of the economic regions within our country. CZs will help.



My open-source code is available on GitHub: https://github.com/jevongibb/Code/tree/master/CommutingZones. Please submit issues and commit improvements.

A more technical discussion (while trying to avoid pedanticism) -

This picture gives a good comparison between different region definitions. I have never met a person who felt defined by their Economic Area (the biggest, unshaded regions below). Metro- and Mico-politan Statistical Areas seem better, but they leave out much of the country.



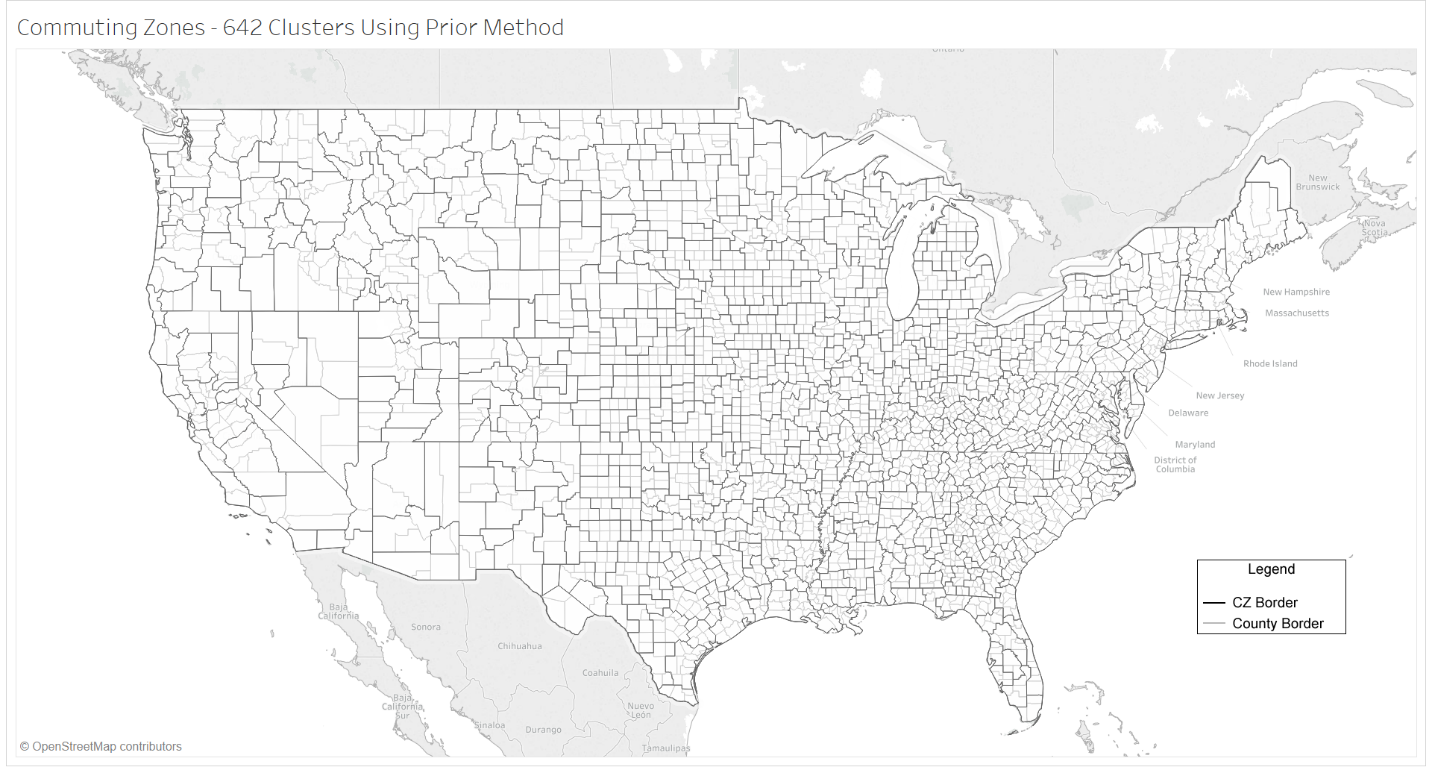
This map shows CZ definitions from year 2000 by the United States Department of Agriculture Economic Research Service.



The USDAERS and Tolbert & Sizer (T&S) methodologies generate around 700-750 CZs.

I attempted to replicate Tolbert & Sizer’s (T&S) methodology using [this paper](https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1044&context=ldi) from Foote, Kutzbach, and Vilhuber. See my R Script on GitHub for details. Of note, T&S cut the national data into multiple regions due to limitations in computing power. We don’t have those limitations anymore. For a similar reason – I assume - they did not seek to optimize the number of clusters. Instead, they used a “rule of thumb”: “as a rule of thumb, a normalized average distance of 0.98 was considered sufficient distance between sets of counties to treat them as separate [Labor Market Areas].”

If you cut my cluster data at .98, you get 642 CZs:



I think a better methodology is to optimize the number of clusters, rather than accepting the rule of thumb. Because CZs use non-Euclidean distance, we cannot use the elbow method. We can, though, use the [Average Silhouette Method](https://www.sciencedirect.com/science/article/pii/0377042787901257?via%3Dihub), which generate an “average silhouette width (that) provides an evaluation of clustering validity, and might be used to select an ‘appropriate’ number of clusters.”

